Indian Internet Journal of Forensic Medicine & Toxicology
ISSN 0973-1970, Vol.7, No.1 (2009)

Genetic Polymorphism at 15 STR Loci in Delhi's (India) population

Anupuma Raina[*], Bhuvnesh Yadav[*], K.V. Bhat[**] Tirath D. Dogra[*], [*]DNA Fingerprinting Laboratory, Department of Forensic Medicine, All India Institute of Medical Sciences, New Delhi, India, [**]National Research Centre for DNA Fingerprinting, National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi, India

Abstract

The migration and genetic drift influence the population structures, which necessitate the genetic polymorphism studies of the population subgroups. Delhi, being capital of India, has mixed population group of the individuals migrated from different parts of the country. Keeping this in view, the allelic database of Delhi population is generated here using 15 STR markers to study the impact of these natural factors (migration, genetic drift, selection etc.) on this population.

For this purpose, the allelic frequencies for 15 STR loci were estimated and observed to be on higher side. Further, the study includes the calculation of observed and expected heterozygosity. Moreover, the genetic variation statistics shows that there is out breeding in this population and also the effective number of alleles is higher. The observed results indicate that all the 15 STR loci are highly polymorphic in this population which indicates that there is high degree of diversity. Also, it signifies that this is a random mating population. Moreover, it is observed that four loci (D3S1358, TH01, D13S317, and D16S539) are undergoing some sort of selection.

Keywords: Short tandem repeat (STR); Indian population data; AmpF*I*STR® Identifiler™ kit, Genetic diversity

Introduction

Indian population is a combination of diverse ethnic, linguistic and geographical groups. The complete genetic data of this population is not yet available, as it is still under study, using the STR markers at various loci.

Forensic DNA Typing is widely used to identify an individual for the purpose of medico-legal cases (Victorian Parliament Law Reform Committee 2002) all over the globe. The usage of

For Correspondence :
Dr. Anupuma Raina,
Room No. 313, Department of Forensic Medicine, All India Institute of Medical Sciences, Delhi, India

population genetic findings and technologies has improved, particularly for the identification of an individual, which is a breakthrough in the field of molecular biology. Forensic cases are analyzed more rapidly, and DNA databanks are generated faster than a decade ago by performing genotyping of short tandem repeats (STR) by multiplex PCR amplification, which is presently one of the preferred methods for forensic casework and other applications. DNA typing is entirely based on the 0.1% difference in human genome that makes every individual unique. This 0.1% includes STRs, VNTRs and SNPs (Butler 2001). The polymorphism at STR loci provide useful marker for purpose forensic analysis (Reddy *et-al*. 2005). It is known, that many factors affect the polymorphism pattern like mutation, selection, migration and admixture. In addition to these, other factors responsible for polymorphism are size, density, distribution of population and marriage rules (Kashyap *et-al*. 2004). The variation at genetic level is necessary to allow organisms to adapt ever-changing environment; also, these variations make every population a unique group. It is possible, that populations living in a close geographical proximity are more likely to exchange their genes, thereby exhibiting genetic similarity, despite the fact that these populations do not belong to the same socio-cultural level. Thus, this pilot study was conducted to check whether this population follows Hardy-Weinberg equilibrium or not, and also, kind of breeding and gene flow in this population. To study the statistical differences among the local population is an important line of attack on the evolutionary studies. As such population differentiation can only be the first step towards speciation in the sense of splitting the species.

Material and Methods

Population

Delhi being a metropolitan city has its inhabitants from various states of the country, and possibly from other countries. The blood samples were collected randomly from the mortuary, Department of Forensic Medicine, AIIMS, and from voluntary donor residents of Delhi. Samples were collected from unrelated individuals; each volunteer giving 100 ml of blood for the study.

Methodology

DNA was isolated from these samples following standard Phenol-Chloroform extraction protocol (Sambrook *et al.*1989). The extracted DNA was quantified on 0.8% agarose gel, visualized on Alpha Imager 3400™ Imaging System and the quality of the DNA was checked by UV Visible Spectrophotometer (Thermo Spectronic, Unichem Ltd). The samples were amplified using AmpF*l*STR® Identifiler™ PCR amplification kit (ABI, Foster City, CA), for 16 STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, Amelogenin, D5S818 and FGA) simultaneously. The PCR amplification was performed using Peltier Thermal Cycler (PTC 200, MJ Research Inc.). Further, fragment separation of amplified products was performed by capillary electrophoresis using ABI 3130 automated Genetic Analyzer. The raw data thus obtained were analyzed, to obtain the allelic designations with data collection software (Genemapper® *ID* software (version 3.2)). The allelic data obtained were statistically analyzed with Arlequin software (v 2.0). The allele frequencies representing repeat numbers were computed to obtain observed and expected heterozygosity

and its standard deviations (Nei and Roychoudhury 1974). The possible divergence from Hardy–Weinberg expectation (HWE) was tested by calculating the unbiased estimate of the expected homozygote/heterozygote frequencies. Further, Ewens-Watterson test for neutrality, ÷2 test and F-statistics were performed to get the complete population information.

Results and Discussion

The overall allelic frequency data, summary of heterozygosity statistics for all loci (observed and expected heterozygosity), and parameters of testing for Hardy-Weinberg equilibrium are given in Table1. The frequencies of most common alleles are highlighted in this table. Calculation of ÷2 and its associated degree of freedom was carried out in order to assess goodness of fit (Hartle and Clark 1997). In Delhi population, for most of the loci, probability is greater than 0.05 indicating that the genotypic frequencies are in Hardy Weinberg proportion for the respective locus. For a few loci such as D21S11, CSF1PO, D7S820, D2S1338, TH01 and FGA, probability is less than 0.05, which indicates, that the deviation from Hardy Weinberg equilibrium is not due to the chance alone, but also on account of some natural factors like migration, selection etc, which are causing the deviation from the Hardy Weinberg model. The heterozygosity statistics shows that all the loci (except D2S1338, TH01, D13S317 and D21S11) have greater observed heterozygosity than expected. Also, the average heterozygosity is high for almost all the loci, suggesting the higher polymorphism, hence, this marker set is highly informative amongst the selected population. Also, the same results have been reported in previous studies for other populations [Bamshad *et-al.* 2001; Ashma and Kashyap 2002b; 2002a; Gaikwad and Kashyap 2002; Das K. 1996; Kashyap *et al* 2002; Rajkumar and Kashyap 2002; Sarkar and Kashyap 2002). On account of high diversity of population, this set of STR can be effectively used for the forensic purposes like paternity testing and individual identification, genetic mapping etc (Gill *et al* 1985, Hammond *et al*. 1994, Gill *et-al*. 1995). The allelic richness of this population is 7.63. The most common alleles for this population are given in figure 1.

Genetic Variation Statistics for all Loci calculated by Nei (Nei 1987) method gives the number (Ne) of equally frequent allele that would be required to produce same homozygosity as observed in the actual population (Table 2). The effective numbers of alleles are sufficiently high, suggesting that the population is highly diverse. Also, the Shannon's informative index is greater, supporting the diversity views. Wright's (1978) fixation index (F_{IS}) as a measure of heterozygote deficiency or excess is not much significant in this population (Table 3). As the F_{IS} value is less than 0.05, indicating little genetic differentiation among the loci [Nei M. 1987, Wright S. 1984]. In this population, D2S1338 is the only locus showing moderate genetic differentiation. For all other loci, no significant genetic differentiation was observed. The observations show that no allele is fixed for any loci in Delhi population. This could be due to the fact that single population is selected for the study and the heterozygosity is so high. F statistics was used to quantify the inbreeding effect of population substructures. As F_{IT} (inbreeding coefficient) values were observed to be negative at many loci, it signifies that the population does not indicate going into any inbreeding at all, and it is totally outgoing for the purpose of mating. It can probably be due to the fact that population is a cluster of the people migrated from the different parts of the India and don't inbreed with each other. The values of fixation index (F_{IS}) and inbreeding coefficient (F_{IT}) were

Table 1. Allelic frequencies of Delhi's Population

Allele/ Locus	D8 S1179	D21 S11	D7 S820	CSF 1PO	D3 S1358	TH 01	D13 S317	D16 S539	D2 S1338	D19 S433	VWA	TPOX	D18 S51	Aml	D5 S818	FGA
6						0.26										
6.3																
7			0.02			0.17										
7.3																
8	0.03		**0.27**	0.02		0.15	0.19	0.06				**0.46**			0.01	
9	0.01		0.08	0.03		**0.27**	0.14	0.15				0.10			0.05	
9.2																
9.3						0.15										
9.5										0.01						
10	0.16		0.20	0.28		0.02	0.06	0.08				0.10	0.02		0.11	
11	0.01		0.22	0.23			0.21	**0.26**		0.01		0.33	0.03		0.26	
11.2	0.01															
12	0.13		0.19	**0.34**		0.01	**0.28**	**0.26**	0.01	0.08		0.02	0.07		**0.32**	
12.2									0.01	0.01						
13	0.19		0.03	0.10			0.12	0.17	0.01	**0.26**			0.13		0.26	
13.2																
14	0.19		0.01	0.01	0.10		0.02	0.04								
14.2										0.08						
15	**0.21**			0.01	**0.29**				0.03	0.23	0.06		**0.17**		0.01	
15.2									0.01	0.07						
16	0.05				0.26				0.01	0.04	**0.20**		0.15			
16.2									0.01	0.03						
17	0.01				0.27				0.07	0.01	0.34		0.08			
18					0.09				**0.20**		0.15		0.04			0.01
18.5									0.01							
19					0.01				0.19		0.11		0.07			0.07
20									0.14		0.01		0.02			0.10
20.3													0.01			
21									0.04				0.02			**0.20**
22									0.10				0.01			0.10

22.2																0.01
23									0.10							0.16
23.2																0.01
23.5									0.01							
23.8																0.01
24		0.01							0.06							0.18
24.2											0.01					
24.8																0.01
25									0.04							0.09
25.2																
26									0.01							0.08
26.1																0.01
27		0.01														0.01
28		0.18														
29		**0.20**														
30		0.16														
30.2		0.03														
31		0.06														
31.2		0.10														
32		0.02														
32.2		0.19														
33		0.01														
33.2		0.05														
34.2		0.01														
35		0.01														
X														**0.71**		
Y														0.29		
HO	0.9000	0.8400	0.7900	0.7800	0.8200	0.7500	0.8700	0.8000	0.8000	0.8600	0.8100	0.7400	0.8600	0.5800	0.7800	0.8900
HE	0.8451	0.8570	0.8040	0.7493	0.7656	0.7977	0.8141	0.8111	0.8822	0.8315	0.7931	0.6659	0.8682	0.4139	0.7583	0.8726
Avg Hetr	0.8445	0.8719	0.8013	0.7624	0.7791	0.8166	0.8020	0.8079	0.8743	0.8586	0.8003	0.6674	0.8758	0.4703	0.7559	0.8836
c2	54.028	313.758	232.907	57.031	7.529	35.274	18.219	16.331	766.392	64.230	19.849	16.562	98.959	16.347	14.401	478.965
DF	55	91	28	28	15	21	21	21	210	78	28	10	91	1	21	105
P	0.512	0.000	0.000	0.001	0.941	0.026	0.635	0.751	0.000	0.869	0.870	0.085	0.267	0.000	0.852	0.000

HO - Observed heterozygosity HE - Expected heterozygosity
Ave_Het - Average Heterozygosity α2 - chi-square
DF - degree of freedom P - probability
Highest frequencies are indicated with the bold letters.
Average heterozygosity is sufficiently high except TPOX locus.

Table 2. Genetic variation statistics

	Genetic Variation Statistics			
Locus	Sample Size	na*	ne*	I*
D8S1179	200	11	6.29	1.97
D21S11	200	14	6.79	2.09
D7S820	200	8	5	1.73
CSF1PO	200	8	3.9	1.52
D3S1358	200	6	4.2	1.52
THO1	200	7	4.85	1.65
D13S317	200	7	5.26	1.76
D16S539	200	7	5.18	1.76
D2S1338	200	21	8.18	2.36
D19S433	200	13	5.79	1.99
VWA	200	8	4.74	1.7
TPOX	200	5	2.96	1.25
D18S51	200	14	7.34	2.2
Aml	200	2	1.7	0.6
D5S818	200	7	4.07	1.52
FGA	200	15	7.6	2.18
Mean	200	9.6	5.24	1.74
St. Dev		4.75	1.73	0.43

* Na = Observed number of Alleles* Ne = Effective number of Alleles; I = Shannon's Information index;Effective population size is significantly high in this population

equal and not much significant in this case. It gives an idea that this population is uniform and has no subgroups.

The impact of different alleles on the loci studied was calculated by the genetic variation statistics and it was observed that Delhi population has very high observed as well as effective number of alleles (Ne). Some loci like D8S1179 (11), D21S11 (14), D2S1338 (21), D19S433 (13), D18S51 (14), VWA (8), CSF1PO (8), D7S820 (8) had higher observed (Na) and expected number of alleles (Ne) indicating the highly polymorphic nature of these loci and signified that the

Table 3. Wright's (1978) fixation index (F_{IS}) -a measure of heterozygote deficiency or excess

Allele/ Locus	D8 S1179	D21 S11	D7 S820	CSF 1PO	D3 S1358	TH 01	D13 S317	D16 S539	D2 S1338	D19 S433	VWA	TPOX	D18 S51	Aml	D5 S818	FGA
6						-0.03										
6.3																
7			-0.02			-0.20										
7.3																
8	-0.03		0.07	-0.02		**0.40**	-0.16	-0.06				-0.09			-0.01	
9	-0.01		-0.08	-0.03		**0.10**	-0.07	-0.02				-0.11			-0.05	
9.3						**0.06**										
9.5										-0.01						
10	**0.05**		0.01	-0.13		-0.02	**0.13**	-0.09				-0.11	-0.02		-0.12	
11	-0.05		-0.10	-0.07			0.04	**0.13**		-0.01		-0.16	**0.31**		0.01	
11.2	-0.01															
12	-0.05		**0.22**	-0.03		-0.01	-0.18	-0.04	-0.01	0.05		-0.02	-0.07		-0.14	
12.2									-0.01	-0.01						
13	-0.10		-0.03	**0.13**			-0.04	0.02	-0.01	-0.03			**0.13**		**0.08**	
13.2									-0.01	-0.02						
14	-0.04		**1.00**	-0.01	-0.11		-0.02	-0.04	-0.01	0.03	-0.08		0.04			
14.2										-0.08						
15	-0.20			-0.01	-0.07				**0.31**	-0.13	-0.06		**0.08**		-0.01	
15.2									-0.01	-0.07						
16	-0.05				-0.09				-0.01	-0.04	-0.13		-0.17			
16.2									-0.01	-0.03						
17	-0.01				-0.05				0.26	-0.01	**0.06**		-0.09			
18					-0.09				0.08		-0.01		-0.04			-0.01
18.5									-0.01							
19					-0.01				0.04		-0.01		-0.07			-0.07
20									0.17		-0.01		-0.02			**0.11**
20.3													-0.01			
21									0.26				-0.02			-0.05
22									0.01				-0.01			-0.11
22.2																-0.01

23									0.11												0.05
23.2																					-0.01
23.5									-0.01												
23.8																					-0.01
24				-0.01					-0.06												-0.02
24.2								-0.01													
24.8																					-0.01
25									-0.04												-0.10
26									-0.01												-0.08
26.1																					-0.01
27				-0.01																	-0.01
27.2																					
28				**0.20**																	
29				**0.14**																	
30				-0.19																	
30.2				-0.03																	
31				**0.11**																	
31.2				-0.11																	
32				-0.02																	
32.2				-0.10																	
33				**1.00**																	
33.2				-0.05																	
34																					
34.2				-0.01																	
35				-0.01																	
X																			0.41		
Y																			0.41		
Total	-0.07	0.01			0.01	-0.05	-0.08	0.06	-0.07	0.01	0.09	-0.04	-0.03	-0.12	0.00	-0.41	-0.03	-0.03			

population is breeding randomly. Only the Amelogenin have only two alleles and hence, have less effective and observed number of alleles.

The Ewens-Watterson test for neutrality is the statistics of natural selection (Table 4). The test for neutrality has thrown up some significant results. It shows that observed frequencies of some loci (D3S1358- 0.2382, THO1- 0.2062, D13S317- 0.1900, D16S539- 0.1930) are not within the upper and lower 95 limits which indicates that this population is undergoing some selection that may be towards positive or negative side of the evolution. These loci may be linked with some fitness related traits, which is under evolutionary process. The data also verifies that the deviation from HWE at D21S11, CSF1PO, D7S820, D2S1338 and FGA is due to migration. In the Ewens-Watterson neutrality test as the values of F from the null hypothesis were larger than the observed F, in this population the observed frequencies in neutrality test are less than

Table 4. Overall Ewens- Watterson Test for Neutrality

Locus	n	k	Obs. F	Min F	Max F	Mean*	SE*	L95*	U95*
D8S1179	200	11	0.1591	0.0909	0.905	0.287	0.0104	0.1553	0.5458
D21S11	200	14	0.1473	0.0714	0.8784	0.2323	0.0074	0.1293	0.4648
D7S820	200	8	0.2	0.125	0.9325	0.3779	0.0179	0.1995	0.6984
CSF1PO	200	8	0.2544	0.125	0.9325	0.3821	0.02	0.1939	0.7288
D3S1358	200	6	**0.2382**	0.1667	0.9513	0.4667	0.0259	0.2384	0.8396
THO1	200	7	**0.2062**	0.1429	0.9418	0.4148	0.202	0.2238	0.7461
D13S317	200	7	**0.19**	0.1429	0.9418	0.4236	0.0232	0.2241	0.778
D16S539	200	7	**0.193**	0.1429	0.9418	0.4152	0.0222	0.2188	0.7863
D2S1338	200	21	0.122	0.0476	0.82	0.149	0.0026	0.0869	0.2751
D19S433	200	13	0.1727	0.0769	0.8872	0.2449	0.0076	0.1358	0.4688
VWA	200	8	0.2109	0.125	0.9325	0.3791	0.0185	0.2016	0.7205
TPOX	200	5	0.3375	0.2	0.9608	0.534	0.031	0.2814	0.9036
D18S51	200	14	0.1362	0.0714	0.8784	0.2302	0.0068	0.1274	0.4343
Aml	200	2	0.5882	0.5	0.99	0.8339	0.0281	0.5032	0.9901
D5S818	200	7	0.2455	0.1429	0.9418	0.4169	0.0212	0.2154	0.7647
FGA	200	15	0.1317	0.667	0.8698	0.213	0.0056	0.1231	0.4067

n - Number of alleles Obs F - Observed frequencies
Min F - Minimum Frequency Max F - Maximum frequency

the values of null hypothesis, it signifies that the common allele is more common than was expected (Manly 1985).

There is an excess of heterozygotes in the total population, as very little or no inbreeding is going on select population. The possible reason for this result, may be that the constituent population are migrating from different parts of India, and these are probably migrating from their permanent home places to settle down in Delhi, and there that there is no inbreeding in the migrant population residing in Delhi. The original population of Delhi is very small and most of the present inhabitants are migrants from all over India.

Acknowledgements

We gratefully acknowledge the contribution of all the voluntary donors who donated their blood to make the study possible. We also thank the mortuary staff for their unbridled co-operation in blood collection from the unclaimed bodies. We extend a special thank to University Grant Commission and All India Institute of Medical Sciences for funding the study.

References

1. Ashma R. and Kashyap V. K. 2002a Genetic study of 15 important STR loci among four major ethnic groups of Bihar, India. J. Forensic Sci. 47, 1139–1142.
2. Ashma R. and Kashyap V. K. 2002b Genetic polymorphism at 15 STR loci among three important sup-populations of Bihar, India. Forensic Sci. Int. 130, 58–62.
3. Bamshad M., Kivisild T., Watkins W. S., Dixon M. E., Ricker C. E., Rao B. B. et al. 2001 Genetic evidence on the origins of Indian caste populations. Genome Res. 11, 994-1004.
4. Butler J. M. 2001 Forensic DNA Typing. 2nd ed. Academic Press, 12-13.
5. Das K, Malhotra K. C.,Mukherjee B. N., Walter H., Majumder P. P. and Papiha S. S. 1996 Population structure and genetic differentiation among 16 tribal populations of Central India. Hum. Bio. 68(5), 679-705.
6. Gaikwad S. and Kashyap V. K. 2002 Polymorphism at fifteen hyper variable microsatellite loci in four populations of Maharashtra, India. Forensic Sci. Int. 126(3), 267–271.
7. Gill P., Jeffreys A. J. and Werrett D. J. 1985 Forensic applications of DNA fingerprints. Nature 318, 577 – 579
8. Gill P., Kimpton C. P., Urquhart A., Oldroyd N., Millican E. S., Watson S. K. et-al. 1995 Automated short tandem repeat (STR) analysis in forensic casework- a strategy for the future. Electrophoresis 16, 1543-1552.
9. Hammond H. A., Jin L., Zhong Y., Caskey C. T. and Chakarborty R. 1994 Evaluation of 13 Short Tandem Repeat loci for use in personal identification applications. Am. J. Hum. Genet. 55, 175-189.
10. Hartle D. L. and Clark A. G. 1997 Principle of Population Genetics. 81-82.
11. Inquiry into Forensic Sampling and DNA Databases: Background / Issue Paper. 2002 Victorian Parliament Law Reform Committee.

12. Kashyap V. K., Ashma R., Gaikwad S., Sarkar B. N. and Trivedi R. 2004 Deciphering diversity in populations of various linguistic and ethnic affiliations of different geographical regions of India: Analysis based on 15 microsatellite markers. J. Genet. 83(1), 49-63.
13. Kashyap V. K., Guha S. and Trivedi R. 2002 Concordance study on 15 STR loci in three major populations of Himalayan State Sikkim. J. Forensic Sci. 47(5), 1163–1167.
14. Manly B. F. J. 1985 The statistics of natural selection on animal populations. Chapman and Hall, London. 272-282.
15. Nei M. 1987 Molecular Evolutionary Genetics. Columbia University Press, New York. 159 –164.
16. Nei M. 1987 Molecular Evolutionary Genetics. Columbia University Press, New York. 176 –187.
17. Nei M., Roychoudhury A.K. 1974 Sampling variance of heterozygosity and genetic distance. Genet. 76, 379-390.
18. Rajkumar R. and Kashyap V. K. 2002 Distribution of alleles of fifteen STR loci of the Powerplex 16™ multiplex in four predominant population groups of South India. Forensic Sci. Int. 126, 175–179.
19. Reddy B. M. Naidu V. M., Madhavi V. K., Thangaraj K., Langstieh B. T., Venkataramana P. et-al. 2005 STR data for the Amp Fl STR Profiler Plus loci among 27 populations of different social hierarchy from southern part of Andhra Pradesh, India. Forensic Sci. Int. 149,81-97
20. Sambrook J., Fritsch E. F. and Maniatis T. 1989 Molecular cloning: a laboratory manual, 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
21. Sarkar N. and Kashyap V. K. 2002 Genetic diversity at two pentanucleotide and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of Central India. Forensic Sci. Int. 128(3), 196–201.
22. Wright S. 1984 Evolution and the genetics of population (Variability within and among natural populations). Vol 4. University of Chicago Press, Chicago.